

BINUS University

Academic Career: <i>Undergraduate / Master / Doctoral *)</i>		Class Program: <i>International/Regular/Smart Program/Global Class*)</i>	
<input type="checkbox"/> Mid Exam <input checked="" type="checkbox"/> Final Exam <input type="checkbox"/> Short Term Exam <input type="checkbox"/> Others Exam : _____		Term : Odd /Even/ Short *)	
<input checked="" type="checkbox"/> Kemanggisan <input checked="" type="checkbox"/> Alam Sutera <input type="checkbox"/> Bekasi <input type="checkbox"/> Senayan <input type="checkbox"/> Bandung <input type="checkbox"/> Malang		Academic Year : 2019 / 2020	
Faculty / Dept. : School of Computer Science		Deadline	Day / Date : Senin / 29 Jun 2020
			Time : 17:00
Code - Course : COMP6577 – Machine Learning		Class : All classes	
Lecturer : Team		Exam Type : Online	
*) <i>Strikethrough the unnecessary items</i>			
<i>The penalty for CHEATING is DROP OUT!!!</i>			

NOTE:

1. Soal ujian berupa kasus sehingga **jawaban Anda sudah pasti tidak akan sama** dengan jawaban teman Anda.
2. Jawaban ditulis dengan bahasa pemrograman **Python** dan dikumpulkan dalam 1 dokumen **Python Notebook**, beserta dengan **analisa dari pertanyaan-pertanyaan** yang diberikan.
3. Jawaban yang dikumpulkan sebaiknya juga berisi mengenai **metode-metode yang telah Anda lakukan sebelumnya**. (Tidak hanya hasil akhir)
4. Nama dokumen *Python Notebook* yang dikumpulkan berupa **NIM.ipynb**
5. Jawaban ujian Anda harus dikumpulkan dalam bentuk *dokumen* yang dikompres dengan format nama dokumen: **NIM.zip**
6. **Ujian akan di-NOL-kan** apabila terindikasi terjadinya kecurangan. Selain itu, ujian Anda **akan diserahkan kepada pihak jurusan** untuk diproses lebih lanjut.

Verified by,

Henry Ham (D5872) and sent to Department on May 29, 2020

I. Kasus (100%)

- Soal ujian ini akan menggunakan “**dataset.csv**” yang berisi mengenai informasi *listing* penyewaan tempat tinggal harian. Dataset dapat diambil pada link berikut: <https://bit.ly/UAS1920-COMP6577-dataset>
- Dataset tersebut mempunyai 30 fitur yang dapat digunakan untuk melakukan analisa metode *machine learning*.
- Anda diwajibkan untuk memilih **12 fitur utama secara random** untuk menjawab pertanyaan-pertanyaan di bawah ini. Fitur utama yang Anda pilih sebaiknya tidak sama dengan variabel yang akan diprediksi.
Note: Probabilitas kombinasi 12 fitur utama seseorang akan sama dengan kombinasi 12 fitur utama orang yang lain sangat kecil.
- Anda perlu membagi “dataset.csv” menjadi dua buah dataset yaitu “**train.csv**” dan “**test.csv**” yang berisi **80% dan 20% dari data keseluruhan**. Perlu dipastikan bahwa dataset *training* dan *testing* tidak boleh mempunyai data yang sama. Anda akan menggunakan dataset *training* pada saat membuat *model machine learning* dan menggunakan dataset *testing* untuk menghitung akurasi.

Berdasarkan informasi di atas, tugas Anda antara lain:

- [20%]** Anda akan melakukan *regression*, *classification*, dan *clustering* dengan menggunakan 12 fitur utama yang telah Anda pilih. Oleh karena itu, **Anda perlu melakukan *feature engineering* dan *feature selection*** untuk setiap metode (soal no. 2-7) di bawah ini.
- [15%]** Lakukan ***logistic regression*** untuk memprediksi harga (*price*) dari masing-masing listing! Jelaskan bagaimana *feature engineering* dan *feature selection* mempengaruhi hasil dari prediksi Anda! Apa metrik yang Anda gunakan untuk menghitung akurasi dari prediksi harga?
- [15%]** Lakukan ***nearest-neighbor classification*** untuk memprediksi nilai review (*review-score-rating*) dari masing-masing listing! Jelaskan bagaimana *feature engineering* dan *feature selection* mempengaruhi hasil dari prediksi Anda! Apa metrik yang Anda gunakan untuk menghitung akurasi dari prediksi harga? Jelaskan bagaimana performa dari metode tersebut dibandingkan metode *classification* lainnya!
- [15%]** Lakukan ***support-vector machine classification*** untuk memprediksi nilai review (*review-score-rating*) dari masing-masing listing! Jelaskan bagaimana *feature engineering* dan *feature selection* mempengaruhi hasil dari prediksi Anda! Apa metrik yang Anda gunakan untuk menghitung akurasi dari prediksi harga? Jelaskan bagaimana performa dari metode tersebut dibandingkan metode *classification* lainnya!
- [15%]** Lakukan ***classification tree*** untuk memprediksi nilai review (*review-score-rating*) dari masing-masing listing! Jelaskan bagaimana *feature engineering* dan *feature selection* mempengaruhi hasil dari prediksi Anda! Apa metrik yang Anda gunakan untuk menghitung akurasi dari prediksi harga? Jelaskan bagaimana performa dari metode tersebut dibandingkan metode *classification* lainnya!
- [10%]** Jelaskan bagaimana ***Gaussian Mixture Model*** menampilkan persebaran data dari fitur-fitur utama yang Anda pilih!
- [10%]** Jelaskan bagaimana ***K-Means Clustering*** menampilkan persebaran data dari fitur-fitur utama yang Anda pilih!

-- Selamat Mengerjakan --

Verified by,

Henry Ham (D5872) and sent to Department on May 29, 2020